

WO9964630

Publication Title:

GENOTYPING THE HUMAN PHENOL SULFOTRANSFERASE 2 GENE (STP2)

Abstract:

Abstract of WO9964630

Genetic polymorphisms are identified in the human STP2 gene that alter STP2-dependent drug metabolism. Nucleic acids comprising the polymorphic sequences are used to screen patients for altered metabolism for STP2 substrates, potential drug-drug interactions, and adverse/side effects, as well as diseases that result from environmental or occupational exposure to toxins. The nucleic acids are used to establish animal, cell and in vitro models for drug metabolism. Data supplied from the esp@cenet database - Worldwide

Courtesy of <http://v3.espacenet.com>

This Patent PDF Generated by Patent Fetcher(R), a service of Stroke of Color, Inc.

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68		A1	(11) International Publication Number: WO 99/64630 (43) International Publication Date: 16 December 1999 (16.12.99)
(21) International Application Number: PCT/US99/13094		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 9 June 1999 (09.06.99)		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(30) Priority Data: 60/088,710 10 June 1998 (10.06.98) US			
(71) Applicant: AXYS PHARMACEUTICALS, INC. [US/US]; 180 Kimball Way, South San Francisco, CA 94080 (US).			
(72) Inventors: GUIDA, Marco; 3899 Nobel Drive #1301, San Diego, CA 92122 (US). KURTH, Janice; 13044 Walking Path Place, San Diego, CA 92130 (US).			
(74) Agent: SHERWOOD, Pamela, J.; Bozicevic, Field & Francis LLP, Suite 200, 285 Hamilton Avenue, Palo Alto, CA 94301 (US).			
(54) Title: GENOTYPING THE HUMAN PHENOL SULFOTRANSFERASE 2 GENE (STP2)			
(57) Abstract			
Genetic polymorphisms are identified in the human STP2 gene that alter STP2-dependent drug metabolism. Nucleic acids comprising the polymorphic sequences are used to screen patients for altered metabolism for STP2 substrates, potential drug-drug interactions, and adverse/side effects, as well as diseases that result from environmental or occupational exposure to toxins. The nucleic acids are used to establish animal, cell and <i>in vitro</i> models for drug metabolism.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

GENOTYPING THE HUMAN PHENOL SULFOTRANSFERASE 2 GENE (STP2)**INTRODUCTION**

Sulfonation is an important pathway in the biotransformation of many drugs, 5 xenobiotics, neurotransmitters, and steroid hormones. Many of the sulfonation reactions for pharmacologic agents are performed by a group of enzymes known as phenol transferases. The phenol sulfotransferase gene family consists to three members located on chromosome 16. A single gene (STM) encodes the thermolabile monoamine-metabolizing form. Two thermostable phenol-metabolizing enzymes are encoded by STP1 and STP2. Substrates for 10 STP1 and STP2 include minoxidil, acetaminophen, and para-nitrophenol. Alterations in phenol sulfotransferase activity have been correlated with individual variation in sulfonation of acetaminophen (Reiter and Weinshilboum (1982) Clin. Pharm.) and predisposition to diet-induced migraine headaches.

The STP2 gene spans approximately 5.1 kb and contains nine exons that range in 15 length from 74 to 347 bp. Exons IA and IB are noncoding and represent two different cDNA 5'-untranslated region sequences. The two apparent 5'-flanking regions of the STP2 gene contain no canonical TATA boxes, but do contain CCAAT elements. STP2 has been localized to human chromosome 16.

Since rates of metabolism of drugs, toxins, etc. can depend on the amounts and kinds 20 of phenol sulfotransferase in tissues, variation in biological response may be determined by the profile of expression of phenol sulfotransferases in each person. Analysis of genetic polymorphisms that lead to altered expression and/or enzyme activity are therefore of interest.

SUMMARY OF THE INVENTION

25 Genetic sequence polymorphisms are identified in the STP2 gene. Nucleic acids comprising the polymorphic sequences are used in screening assays, and for genotyping individuals. The genotyping information is used to predict an individuals' rate of metabolism for STP2 substrates, potential drug-drug interactions, and adverse/side effects. Specific polynucleotides include the polymorphic STP2 sequences set forth in SEQ ID NOs:63-100.

30 The nucleic acid sequences of the invention may be provided as probes for detection of STP2 locus polymorphisms, where the probe comprises a polymorphic sequence of SEQ ID NOs:63-110. The sequences may further be utilized as an array of oligonucleotides comprising two or more probes for detection of STP2 locus polymorphisms.

35 Another aspect of the invention provides a method for detecting in an individual a polymorphism in STP2 metabolism of a substrate, where the method comprises analyzing the genome of the individual for the presence of at least one STP2 polymorphism; wherein the

presence of the predisposing polymorphism is indicative of an alteration in STP2 expression or activity. The analyzing step of the method may be accomplished by detection of specific binding between the individual's genomic DNA with an array of oligonucleotides comprising STP2 locus polymorphic sequences. In other embodiments, the alteration in STP2 expression or activity is tissue specific, or is in response to a STP2 modifier that induces or inhibits STP2 expression.

DATABASE REFERENCES FOR NUCLEOTIDE SEQUENCES

Genbank accession no. U34804 provides the sequence of the STP2 gene.

10

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

STP2 Reference Sequences. SEQ ID NO: 1 lists the sequence of the reference STP2 gene. The exons are as follows: exon 1A (nt 2591-2664); exon 1B (nt 3180-3526); exon 2 (nt 3726-3877); exon 3 (nt 3985-4110); exon 4 (nt 4196-4293); exon 5 (nt 6088-6214); exon 6 (6310-6404); exon 7 (nt 7214-7394); exon 8 (nt 7517-7712). The mRNA sequence is set forth in SEQ ID NO:2, and the encoded amino acid sequence in SEQ ID NO:3.

Primers. The PCR primers for amplification of polymorphic sequences are set forth as SEQ ID NOs:4-17. The primers used in sequencing isolated polymorphic sequences are presented as SEQ ID NOs:18-46. The primers used in Taqman assays are listed as SEQ ID NO:47-62.

Polymorphisms. Polymorphic sequences of STP2 are presented as SEQ ID NOs:63-110.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

25 Pharmacogenetics is the linkage between an individual's genotype and that individual's ability to metabolize or react to a therapeutic agent. Differences in metabolism or target sensitivity can lead to severe toxicity or therapeutic failure by altering the relation between bioactive dose and blood concentration of the drug. Relationships between polymorphisms in metabolic enzymes or drug targets and both response and toxicity can be used to optimize therapeutic dose administration.

30 Genetic polymorphisms are identified in the STP2 gene. Nucleic acids comprising the polymorphic sequences are used to screen patients for altered metabolism for STP2 substrates, potential drug-drug interactions, and adverse/side effects, as well as diseases that result from environmental or occupational exposure to toxins. The nucleic acids are used to 35 establish animal, cell culture and *in vitro* cell-free models for drug metabolism.

Definitions

It is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, animal species or genera, constructs, and reagents described, as such may vary. It is also to be understood that the terminology used herein is for the purpose of 5 describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

As used herein the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a construct" includes a plurality of such constructs and reference to "the STP2 nucleic acid" includes reference to 10 one or more nucleic acids and equivalents thereof known to those skilled in the art, and so forth. All technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs unless clearly indicated otherwise.

15 *STP2 reference sequence.* The sequence of the STP2 gene may be accessed through Genbank as previously cited, and is provided in SEQ ID NO:1 and SEQ ID NO:2 (cDNA sequence). The amino acid sequence of STP2 is listed as SEQ ID NO:3. These sequences provide a reference for the polymorphisms of the invention. The nucleotide sequences provided herein differ from the published sequence at certain positions throughout the 20 sequence. Where there is a discrepancy the provided sequence is used as a reference.

The term "wild-type" may be used to refer to the reference coding sequences of STP2, and the term "variant", or "STP2" to refer to the provided variations in the STP2 sequence. Where there is no published form, such as in the intron sequences, the term wild-type may be used to refer to the most commonly found allele. It will be understood by one of skill in the art 25 that the designation as "wild-type" is merely a convenient label for a common allele, and should not be construed as conferring any particular property on that form of the sequence.

30 *STP2 polymorphic sequences.* It has been found that specific sites in the STP2 gene sequence are polymorphic, i.e. within a population, more than one nucleotide (G, A, T, C) is found at a specific position. Polymorphisms may provide functional differences in the genetic sequence, through changes in the encoded polypeptide, changes in mRNA stability, binding of transcriptional and translation factors to the DNA or RNA, and the like. The polymorphisms are also used as single nucleotide polymorphisms to detect association with, or genetic linkage to phenotypic variation in activity and expression of STP2.

SNPs are generally biallelic systems, that is, there are two alleles that an individual may have for any particular marker. SNPs, found approximately every kilobase, offer the potential for generating very high density genetic maps, which will be extremely useful for developing haplotyping systems for genes or regions of interest, and because of the nature of SNPs, they 5 may in fact be the polymorphisms associated with the disease phenotypes under study. The low mutation rate of SNPs also makes them excellent markers for studying complex genetic traits.

Single nucleotide polymorphisms are provided in the STP2 promoter, intron and exon sequences. Table 4 and the corresponding sequence listing provide both forms of each 10 polymorphic sequence. For example, SEQ ID NO:99 and 100 are the alternative forms of a single polymorphic site. The provided sequences also encompass the complementary sequence corresponding to any of the provided polymorphisms.

In order to provide an unambiguous identification of the specific site of a polymorphism, sequences flanking the polymorphic site are shown in Table 4, where the 5' and 3' flanking 15 sequence is non-polymorphic, and the central position, shown in bold, is variable. It will be understood that there is no special significance to the length of non-polymorphic flanking sequence that is included, except to aid in positioning the polymorphism in the genomic sequence. The STP2 exon sequences have been published, and therefore one of each pair of sequences in Table 4 is a publically known sequence.

20 As used herein, the term "STP2 gene" is intended to generically refer to both the wild-type and variant forms of the sequence, unless specifically denoted otherwise. As it is commonly used in the art, the term "gene" is intended to refer to the genomic region encompassing 5' UTR, exons, introns, and 3' UTR. Individual segments may be specifically referred to, e.g. exon 2, intron 5, etc. Combinations of such segments that provide for a 25 complete STP2 protein may be referred to generically as a protein coding sequence.

Nucleic acids of interest comprise the provided STP2^v nucleic acid sequence(s), as set forth in Table 4. Such nucleic acids include short hybridization probes, protein coding sequences, variant forms of STP2 cDNA, segments, e.g. exons, introns, etc., and the like. Methods of producing nucleic acids are well-known in the art, including chemical synthesis, 30 cDNA or genomic cloning, PCR amplification, etc.

For the most part, DNA fragments will be of at least 15 nt, usually at least 20 nt, often at least 50 nt. Such small DNA fragments are useful as primers for PCR, hybridization screening, etc. Larger DNA fragments, i.e. greater than 100 nt are useful for production of the encoded polypeptide, promoter motifs, etc. For use in amplification reactions, such as PCR, 35 a pair of primers will be used. The exact composition of primer sequences is not critical to the

invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art.

The STP2 nucleic acid sequences are isolated and obtained in substantial purity, generally as other than an intact or naturally occurring mammalian chromosome. Usually, the 5 DNA will be obtained substantially free of other nucleic acid sequences that do not include a STP2 sequence or fragment thereof, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", i.e. flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

For screening purposes, hybridization probes of the polymorphic sequences may be 10 used where both forms are present, either in separate reactions, spatially separated on a solid phase matrix, or labeled such that they can be distinguished from each other. Assays may utilize nucleic acids that hybridize to one or more of the described polymorphisms.

An array may include all or a subset of the polymorphisms listed in Table 4. One or both polymorphic forms may be present in the array, for example the polymorphism of SEQ 15 ID NO:37 and 38 may be represented by either, or both, of the listed sequences. Usually such an array will include at least 2 different polymorphic sequences, i.e. polymorphisms located at unique positions within the locus, and may include as many all of the provided polymorphisms. Arrays of interest may further comprise sequences, including polymorphisms, of other genetic sequences, particularly other sequences of interest for pharmacogenetic screening, e.g. STP1; 20 UGT1, UGT2, cytochrome oxidases, etc. The oligonucleotide sequence on the array will usually be at least about 12 nt in length, may be the length of the provided polymorphic sequences, or may extend into the flanking regions to generate fragments of 100 to 200 nt in length. For examples of arrays, see Ramsay (1998) Nat. Biotech. 16:40-44; Hacia et al. (1996) Nature Genetics 14:441-447; Lockhart et al. (1996) Nature Biotechnol. 14:1675-1680; and De 25 Risi et al. (1996) Nature Genetics 14:457-460.

Nucleic acids may be naturally occurring, e.g. DNA or RNA, or may be synthetic 30 analogs, as known in the art. Such analogs may be preferred for use as probes because of superior stability under assay conditions. Modifications in the native structure, including alterations in the backbone, sugars or heterocyclic bases, have been shown to increase intracellular stability and binding affinity. Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O- phosphorothioate, 3'-CH₂-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace 35 the entire ribose phosphodiester backbone with a peptide linkage.

Sugar modifications are also used to enhance stability and affinity. The α -anomer of deoxyribose may be used, where the base is inverted with respect to the natural β -anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O- methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity.

5 Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'- deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5- propynyl-2'- deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

10

STP2 *polypeptides*. A subset of the provided nucleic acid polymorphisms in STP2 exons confer a change in the corresponding amino acid sequence. Using the amino acid sequence provided in SEQ ID NO:3 as a reference, the amino acid polymorphisms of the invention include pro \rightarrow leu, pos. 19; ala \rightarrow val, pos. 136; asn \rightarrow thr, pos. 235; glu \rightarrow lys, pos 282; 15 and a truncated form resulting from a stop codon at exon 5, position 447. Polypeptides comprising at least one of the provided polymorphisms (STP2 v polypeptides) are of interest. The term "STP2 v polypeptides" as used herein includes complete STP2 protein forms, e.g. such splicing variants as known in the art, and fragments thereof, which fragments may comprise short polypeptides, epitopes, functional domains; binding sites; etc.; and including 20 fusions of the subject polypeptides to other proteins or parts thereof. Polypeptides will usually be at least about 8 amino acids in length, more usually at least about 12 amino acids in length, and may be 20 amino acids or longer, up to substantially the complete protein.

The STP2 genetic sequence, including polymorphisms, may be employed for polypeptide synthesis. For expression, an expression cassette may be employed, providing 25 for a transcriptional and translational initiation region, which may be inducible or constitutive, where the coding region is operably linked under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination region. Various transcriptional initiation regions may be employed that are functional in the expression host. The polypeptides may be expressed in prokaryotes or eukaryotes in accordance with 30 conventional ways, depending upon the purpose for expression. Small peptides can also be synthesized in the laboratory.

Substrate. A substrate is a chemical entity that is modified by STP2, usually under normal physiological conditions. Although the duration of drug action tends to be shortened 35 by metabolic transformation, drug metabolism is not "detoxification". Frequently the metabolic

product has greater biologic activity than the drug itself. In some cases the desirable pharmacologic actions are entirely attributable to metabolites, the administered drugs themselves being inert. Likewise, the toxic side effects of some drugs may be due in whole or in part to metabolic products.

5 Substrates of interest may be drugs, xenobiotics, neurotransmitters, steroid hormones, etc. STP2 preferentially catalyzes the sulfonation of 'simple' planar phenols. Substrates include minoxidil, acetaminophen, para-nitrophenol, N-hydroxy 4-aminobiphenyl, etc.

10 *Modifier.* A modifier is a chemical agent that modulates the action of STP2, either through altering its enzymatic activity (enzymatic modifier) or through modulation of expression (expression modifier, e.g., by affecting transcription or translation). In some cases the modifier may also be a substrate. Inhibitors include N-ethylmaleimide; phenylglyoxal; 2,6-dichloro-4-nitrophenol; p-nitrophenol; quercetin and other flavonoids, e.g. fisetin, galangin, myricetin, kaempferol, chrysins, apigenin; and phenols such as curcumin, genistein, ellagic acid. Steroids, 15 e.g. estradiol benzoate, testosterone propionate may affect activity and/or expression.

20 *Pharmacokinetic parameters.* Pharmacokinetic parameters provide fundamental data for designing safe and effective dosage regimens. A drug's volume of distribution, clearance, and the derived parameter, half-life, are particularly important, as they determine the degree of fluctuation between a maximum and minimum plasma concentration during a dosage interval, the magnitude of steady state concentration and the time to reach steady state plasma concentration upon chronic dosing. Parameters derived from *in vivo* drug administration are useful in determining the clinical effect of a particular STP2 genotype.

25 *Expression assay.* An assay to determine the effect of a sequence polymorphism on STP2 expression. Expression assays may be performed in cell-free extracts, or by transforming cells with a suitable vector. Alterations in expression may occur in the basal level that is expressed in one or more cell types, or in the effect that an expression modifier has on the ability of the gene to be inhibited or induced. Expression levels of a variant alleles are 30 compared by various methods known in the art. Methods for determining promoter or enhancer strength include quantitation of the expressed natural protein; insertion of the variant control element into a vector with a reporter gene such as β -galactosidase, luciferase, chloramphenicol acetyltransferase, etc. that provides for convenient quantitation; and the like.

35 Gel shift or electrophoretic mobility shift assay provides a simple and rapid method for detecting DNA-binding proteins (Ausubel, F.M. et al. (1989) In: Current Protocols in Molecular Biology, Vol. 2, John Wiley and Sons, New York). This method has been used widely in the

study of sequence-specific DNA-binding proteins, such as transcription factors. The assay is based on the observation that complexes of protein and DNA migrate through a nondenaturing polyacrylamide gel more slowly than free DNA fragments or double-stranded oligonucleotides. The gel shift assay is performed by incubating a purified protein, or a complex mixture of 5 proteins (such as nuclear or cell extract preparations), with an end-labeled DNA fragment containing the putative protein binding site. The reaction products are then analyzed on a nondenaturing polyacrylamide gel. The specificity of the DNA-binding protein for the putative binding site is established by competition experiments using DNA fragments or oligonucleotides containing a binding site for the protein of interest, or other unrelated DNA 10 sequences.

Expression assays can be used to detect differences in expression of polymorphisms with respect to tissue specificity, expression level, or expression in response to exposure to various substrates, and/or timing of expression during development. For example, since STP2 is expressed in liver, polymorphisms could be evaluated for expression in tissues other than 15 liver, or expression in liver tissue relative to a reference STP2 polypeptide.

Substrate screening assay. Substrate screening assays are used to determine the metabolic activity of a STP2 protein or peptide fragment on a substrate. Many suitable assays are known in the art, including the use of primary or cultured cells, genetically modified cells 20 (e.g., where DNA encoding the STP2 polymorphism to be studied is introduced into the cell within an artificial construct), cell-free systems, e.g. microsomal preparations or recombinantly produced enzymes in a suitable buffer, or in animals, including human clinical trials. Where genetically modified cells are used, since most cell lines do not express STP2 activity (liver cell lines being the exception), introduction of artificial construct for expression of the STP2 25 polymorphism into many human and non-human cell lines does not require additional modification of the host to inactivate endogenous STP2 expression/activity. Clinical trials may monitor serum, urine, etc. levels of the substrate or its metabolite(s).

Typically a candidate substrate is input into the assay system, and the oxidation to a metabolite is measured over time. The choice of detection system is determined by the 30 substrate and the specific assay parameters. Assays are conventionally run, and will include negative and positive controls, varying concentrations of substrate and enzyme, etc. Exemplary assays may be found in the literature, for examples see Chou *et al.* (1995) Carcinogenesis 16:413-417; Walle and Walle (1991) Drug Metab. Dispos. 19:448-453; and Falany *et al.* (1990) Arch. Biochem. Biophys. 278:312-318.

Genotyping: STP2 genotyping is performed by DNA or RNA sequence and/or hybridization analysis of any convenient sample from a patient, e.g. biopsy material, blood sample (serum, plasma, etc.), buccal cell sample, etc. A nucleic acid sample from an individual is analyzed for the presence of polymorphisms in STP2, particularly those that affect the 5 activity or expression of STP2. Specific sequences of interest include any polymorphism that leads to changes in basal expression in one or more tissues, to changes in the modulation of STP2 expression by modifiers, or alterations in STP2 substrate specificity and/or activity.

Linkage Analysis: Diagnostic screening may be performed for polymorphisms that are 10 genetically linked to a phenotypic variant in STP2 activity or expression, particularly through the use of microsatellite markers or single nucleotide polymorphisms (SNP). The microsatellite or SNP polymorphism itself may not phenotypically expressed, but is linked to sequences that result in altered activity or expression. Two polymorphic variants may be in linkage disequilibrium, i.e. where alleles show non-random associations between genes even though 15 individual loci are in Hardy-Weinberg equilibrium.

Linkage analysis may be performed alone, or in combination with direct detection of phenotypically evident polymorphisms. The use of microsatellite markers for genotyping is well documented. For examples, see Mansfield et al. (1994) Genomics 24:225-233; and Ziegler et al. (1992) Genomics 14:1026-1031. The use of SNPs for genotyping is illustrated in Underhill 20 et al. (1996) Proc Natl Acad Sci U S A 93:196-200.

Transgenic animals. The subject nucleic acids can be used to generate genetically modified non-human animals or site specific gene modifications in cell lines. The term "transgenic" is intended to encompass genetically modified animals having a deletion or other 25 knock-out of STP2 gene activity, having an exogenous STP2 gene that is stably transmitted in the host cells, or having an exogenous STP2 promoter operably linked to a reporter gene. Transgenic animals may be made through homologous recombination, where the STP2 locus is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, 30 and the like. Of interest are transgenic mammals, e.g. cows, pigs, goats, horses, etc., and particularly rodents, e.g. rats, mice, etc.

Genetically Modified Cells. Primary or cloned cells and cell lines are modified by the introduction of vectors comprising STP2 gene polymorphisms. The gene may comprise one 35 or more variant sequences, preferably a haplotype of commonly occurring combinations. In one embodiment of the invention, a panel of two or more genetically modified cell lines, each

cell line comprising a STP2 polymorphism, are provided for substrate and/or expression assays. The panel may further comprise cells genetically modified with other genetic sequences, including polymorphisms, particularly other sequences of interest for pharmacogenetic screening, e.g. STP1; UGT1, UGT2, cytochrome oxidases, etc.

5 Vectors useful for introduction of the gene include plasmids and viral vectors, e.g. retroviral-based vectors, adenovirus vectors, etc. that are maintained transiently or stably in mammalian cells. A wide variety of vectors can be employed for transfection and/or integration of the gene into the genome of the cells. Alternatively, micro-injection may be employed, fusion, or the like for introduction of genes into a suitable host cell.

10

Genotyping Methods

The effect of a polymorphism in the STP2 gene sequence on the response to a particular substrate or modifier of STP2 is determined by *in vitro* or *in vivo* assays. Such assays may include monitoring the metabolism of a substrate during clinical trials to determine 15 the STP2 enzymatic activity, specificity or expression level. Generally, *in vitro* assays are useful in determining the direct effect of a particular polymorphism, while clinical studies will also detect an enzyme phenotype that is genetically linked to a polymorphism.

20 The response of an individual to the substrate or modifier can then be predicted by determining the STP2 genotype, with respect to the polymorphism. Where there is a differential distribution of a polymorphism by racial background, guidelines for drug administration can be generally tailored to a particular ethnic group.

25 The basal expression level in different tissue may be determined by analysis of tissue samples from individuals typed for the presence or absence of a specific polymorphism. Any convenient method may be used, e.g. ELISA, RIA, etc. for protein quantitation, northern blot or other hybridization analysis, quantitative RT-PCR, etc. for mRNA quantitation. The tissue specific expression is correlated with the genotype.

30 The alteration of STP2 expression in response to a modifier is determined by administering or combining the candidate modifier with an expression system, e.g. animal, cell, *in vitro* transcription assay, etc. The effect of the modifier on STP2 transcription and/or steady state mRNA levels is determined. As with the basal expression levels, tissue specific interactions are of interest. Correlations are made between the ability of an expression modifier to affect STP2 activity, and the presence of the provided polymorphisms. A panel of different modifiers, cell types, etc. may be screened in order to determine the effect under a number of different conditions.

35 A STP2 polymorphism that results in altered enzyme activity or specificity is determined by a variety of assays known in the art. The enzyme may be tested for metabolism of a

substrate *in vitro*, for example in defined buffer, or in cell or subcellular lysates, where the ability of a substrate to be metabolized by STP2 under physiologic conditions is determined. Where there are not significant issues of toxicity from the substrate or metabolite(s), *in vivo* human trials may be utilized, as previously described.

5 The genotype of an individual is determined with respect to the provided STP2 gene polymorphisms. The genotype is useful for determining the presence of a phenotypically evident polymorphism, and for determining the linkage of a polymorphism to phenotypic change.

A number of methods are available for analyzing nucleic acids for the presence of a 10 specific sequence. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. The nucleic acid may be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis. The use of 15 the polymerase chain reaction is described in Saiki et al. (1985) Science 230:1350-1354, and a review of current techniques may be found in Sambrook et al. Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp.14.2-14.33. Amplification may be used to determine 20 whether a polymorphism is present, by using a primer that is specific for the polymorphism. Alternatively, various methods are known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, for examples see Riley et al. (1990) Nucleic Acids Res 18:2887-2890; and Delahunty et al. (1996) Am J Hum Genet 58:1239-1246.

A detectable label may be included in an amplification reaction. Suitable labels include 25 fluorochromes, e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, e.g. 32P, 35S, 3H; etc. The label may be a 30 two stage system, where the amplified DNA is conjugated to biotin, haptens, etc. having a high affinity binding partner, e.g. avidin, specific antibodies, etc., where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

The sample nucleic acid, e.g. amplified or cloned fragment, is analyzed by one of a 35 number of methods known in the art. The nucleic acid may be sequenced by dideoxy or other methods. Hybridization with the variant sequence may also be used to determine its presence, by Southern blots, dot blots, etc. The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilized on a solid support, as described in U.S.

5,445,934, or in WO95/35505, may also be used as a means of detecting the presence of variant sequences. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), mismatch cleavage detection, and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation

5 as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease (restriction fragment length polymorphism, RFLP), the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

10 In one embodiment of the invention, an array of oligonucleotides are provided, where discrete positions on the array are complementary to one or more of the provided polymorphic sequences, e.g. oligonucleotides of at least 12 nt, frequently 20 nt, or larger, and including the sequence flanking the polymorphic position. Such an array may comprise a series of oligonucleotides, each of which can specifically hybridize to a different polymorphism. For

15 examples of arrays, see Hacia et al. (1996) Nat Genet 14:441-447 and DeRisi et al. (1996) Nat Genet 14:457-460.

20 The genotype information is used to predict the response of the individual to a particular STP2 substrate or modifier. Where an expression modifier inhibits STP2 expression, then drugs that are a STP2 substrate will be metabolized more slowly if the modifier is co-administered. Where an expression modifier induces STP2 expression, a co-administered substrate will typically be metabolized more rapidly. Similarly, changes in STP2 activity will affect the metabolism of an administered drug. The pharmacokinetic effect of the interaction will depend on the metabolite that is produced, e.g. a prodrug is metabolized to an active form, a drug is metabolized to an inactive form, an environmental compound is metabolized to a

25 toxin, etc. Consideration is given to the route of administration, drug-drug interactions, drug dosage, etc.

EXPERIMENTAL

30 The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention, and are not intended to limit the scope of what is regarded as the invention. Efforts have been made to ensure accuracy with respect to the numbers used (e.g. amounts, temperature, concentrations, etc.) but some experimental errors and deviations should be allowed for. Unless otherwise indicated, parts are parts by weight, molecular weight is average molecular

35 weight, temperature is in degrees centigrade; and pressure is at or near atmospheric.

MATERIALS AND METHODS

5 *DNA samples.* Blood specimens from approximately 300 individuals were collected after obtaining informed consent. All samples were stripped of personal identifiers to maintain confidentiality. The only data associated with a given blood sample was gender and self-reported major racial group designations in the United States (Caucasian, Hispanic, African American). Genomic DNA was isolated from these samples using standard techniques. gDNA was either stored as concentrated solutions or stored dried in microtiter plates for future use.

10 *PCR amplifications.* The primers used to amplify the coding regions and the promoter region of the STP2 gene from 200 ng of human gDNA are shown in Table 1. Primers were designed based upon publicly available genomic sequence provided by Her *et al.* (1996) *Genomics* 33:409-420. 100 ng of gDNA from 2 individuals was amplified with the Perkin Elmer GeneAmp PCR kit according to manufacturer's instructions in 100 μ l reactions with Taq Gold DNA polymerase, with two exceptions. Boehringer-Mannheim Expand High Fidelity PCR System kit was used to amplify the promoter region and exon 1A. Magnesium concentrations 15 for each PCR reaction was optimized empirically, and are shown in Table 1.

Table 1. PCR primers and Mg⁺⁺ concentrations.

<u>Region</u>	<u>Forward/ Reverse</u>	<u>SEQ ID</u>	<u>Forward Primer (5' 3')</u>	<u>[Mg⁺⁺]</u>
Promoter	F	4.	CCCAAATAACAGGTGTTCC	2mM
	R	5.	GGAGCAGAGCAAGGATC	
Exon 1A	F	6.	TTCTTCTAGGATCTTCTATCG	2mM
	R	7.	ACTCAGCAAAAGGAGGAT	
Exon 1B	F	8.	TTAGAGATGGGGTCTTCC	2mM
	R	9.	GGGCGAGAGATGTCC	
Exon 2	F	10.	GGAGAGGAGCCTACTGG	2mM
	R	11.	AGTCTGAGGTGAGCAT	
Exons 3&4	F	12.	GCCTCAGTGACTCCCT	3mM
	R	13.	TTTGGAAAGAGACTTATCTGG	
Exons 5&6	F	14.	GCAGGACTTGGCTTT	2mM
	R	15.	GACTCAGGCACAGGAG	
Exons 7&8	F	16.	GACCATCCCAGTCCTT	2mM
	R	17.	CCCCAACGACACAGG	

35 *Thermal cycling was performed in a GeneAmp PCR System 9600 PCR machine (Perkin Elmer) with an initial denaturation step at 95°C for 10 min, followed by 35 cycles of denaturation at 95°C for 30 sec, primer annealing at 60°C for 45 sec, and primer extension at 72°C for 2 min, followed by final extension at 72°C for 5 min, with the following exceptions. 40 cycles were used to amplify exon 1B and to co-amplify exons 7 and 8. Cycling conditions for the promoter region and exon 1A were an initial denaturation at 95°C for 2 min, followed by 40*

cycles of denaturation at 94°C for 30 sec, primer annealing at 60 °C for 45 sec, and primer extension at 68°C for 4 min, followed by a final extension at 68°C for 7 min.

5 *DNA sequencing.* PCR products from 32 individuals, approximately 1/3 from each of the 3 major racial groups (see above), were spin column purified using Microcon-100 columns. Cycle sequencing was performed on the GeneAmp PCR System 9600 PCR machine (Perkin Elmer) using the ABI Prism dRhodamine Terminator Cycle Sequencing Ready Reaction Kit according to the manufacturer's directions. Oligonucleotide primers used for the sequencing reactions are listed in Table 2.

10

Table 2. Sequencing primers.

	<u>Region</u>	<u>Forward/ Reverse</u>	<u>SEQ ID</u>	<u>Forward Primer (5' 3')</u>
15	Promoter (1)	F	18.	TGGAGCCCGTCTTGG
		R	19.	CAGCAGTTCACTTGACC
20	Promoter (2)	F	20.	TGCCACCCCTGCT
		R	21.	AGGCTGCTCCCCCTG
25	Promoter (3)	F	22.	GGGCTCACGCAACC
		R	23.	GCAGGTACTTTCTTCCA
30	Exon 1A (1)	F	24.	TTCTTCTAGGATCTTCTATCG
		R	25.	TTTTGAGGTGTCACTGG
35	Exon 1A (2)	F	26.	CCCACACAAACACCCAC
		R	27.	GCTTCTGGAATGTTGG
40	Exon 1A (3)	R	28.	CGGAAAAAAAAGGAAG
		F	29.	CAATGCTGCCAGA
45	Exon 1B (1)	R	30.	GCTCCACTGAGGAACCT
		F	31.	GGAGAGGAGCCTACTGG
50	Exon 1B (2)	R	32.	TACCACCATCACAAACAGC
		F	33.	CTGAAAGCAAGAAATCCAC
55	Exon 2	R	34.	AGGCTGAGGTGAGCAT
		F	35.	GCGGTGACCTGGAA
60	Exons 3&4	R	36.	TTTGGAAAGAGACTTATCTGG
		F	37.	CTGACTTGCCCCCTACCT
65	Exons 5&6 (1)	R	38.	TAGCCACCACCCCTTA
		F	39.	CCAAAGTGTACCCCTCACC
70	Exons 5&6 (2)	R	40.	AGCCTGCTGCCACA
		F	41.	GACCATCCCAGTCCTT
75	Exons 7&8 (1)	R	42.	CAAACCCCCGTGCT
		F	43.	CTGTGGACCTCTGGTTG
80	Exons 7&8 (2)	R	44.	CACAAATCATTTATTCTGG
		F	45.	CGATGCGGACTATGC
85	Exons 7&8 (3)	R	46.	CCCCAACGACACAGG

Eight μ l sequencing reactions were subjected to 30 cycles at 96°C for 20 sec, 50°C for 20 sec, and 60°C for 4 min, followed by ethanol precipitation. Samples were evaporated to dryness at 50°C for ~15 min and resuspended in 2 μ l of loading buffer (5:1 deionized formamide:50 mM EDTA pH 8.0), heated to 65°C for 5 min, and electrophoresed through 4% 5 polyacrylamide/6M urea gels in an ABI 377 Nucleic Acid Analyzer according to the manufacturer's instructions for sequence determination. All sequences were determined from both the 5' and 3' (sense and antisense) direction. Each sequencing reaction was performed with 2 individuals' DNA pooled together. The 16 electropherograms were analyzed by comparing peak heights, looking for ~25% reduction in peak size and/or presence of extra 10 peaks as an indication of heterozygosity. If polymorphisms were identified, pools were subsequently split and resequenced for confirmation.

15 *Population genotyping.* High-throughput genotyping using TaqMan technology (ABI) was performed using standard techniques (Livak et al. (1995) PCR Methods and Applications 4:357-362) on the samples described above for 3 STP2 polymorphisms. Oligonucleotide PCR primers and probes used for genotyping are shown in Table 3. Polymorphisms for which allele frequencies were determined are marked with an asterisk (*) in Table 4.

Table 3. TaqMan primers and probes.

20	SEQ ID	Description	Primers
47.		STP2-136A primer	GGTGCTGGGTTGAGTCTTCTG
48.		STP2-136Ala probe	CAAAGGATGTGGGGTTCCACTACC
49.		STP2-136B primer	ACACCTCCTCCCTCCATCAAG
50.		STP2-136Val probe	CGAAAGGATGTGGTGGTTCCACTAC
25	51.	STP2-235A primer	GGAGACTGTGGACCTCATGGTGA
	52.	STP2-235Asn probe	TAGTTGGTCATAGGGTCTTCTCATCTCCTT
	53.	STP2-235B primer	CCGGCACCTACCTTCCTCAT
	54.	STP2-235Thr probe	TAGTTGGTCATAGGGTCTTCTCATCTCC
	55.	STP2-282A primer	AGTTTGCTCCCTGCCTTCCT
30	56.	STP2-282Glu probe	CTGCCATCTTCTCCGCATAGTCGG
	57.	STP2-282B primer	GGAACCCCTCTCACAGCTCAGA
	58.	STP2-282Lys probe	TGCCATCTTCTTCGCATAGTCGGC
	59.	STP2-447A primer	GGTGCTGGGTTGAGTCTTCTG
	60.	STP2-DelA447 probe	ATGGCCAAAGTGTACCCCTACCCCTG
35	61.	STP2-447B primer	ACACCTCCTCCCTCCATCAAG
	62.	STP2-InsA447 probe	CATGGCCAAAGTGTAA CC CTCACCC

40 Assay name is given by locus and position. Primer names are abbreviated locus - position and letter designations representing forward (A) and reverse (B) primers. Probes are abbreviated locus-position and 3 letter nucleic acid designations representing the nucleic acid alteration in the coding strand of the genomic DNA. Positions at which probes detect nucleic acid variations are shown in bold.

RESULTS

Eight exons, the promoter region, 3' and 5' untranslated regions from the human STP2 gene were resequenced in 32 individuals representing three major ethnic groups (Caucasian, Hispanic, and African American). The polymorphisms are listed in Table 4.

5 **Table 4.** Newly identified STP2 gene polymorphisms.

	Location;	SEQ ID	Polymorphism Sequence	AA change
	3' end; 99	63.	CCAGCTCCTCA ACTT GCCCTG	
		64.	CCAGCTCCT T ACTTGCCCTG	
10	3' UTR; 7	65.	GTGAGAGGGGTTCCCTGGAGTC	
		66.	GTGAGAGGGGCTCCTGGAGTC	
	Promoter; -603	67.	CATGAAGCTGGGGCTGGCTCC	
		68.	CATGAAGCT G AGGCTGGCTCC	
	Promoter; -833	69.	CTCGTCCCCAGCTTGACCTG	
		70.	CTCGTCCCC A CTTGACCTG	
15	Promoter; -1005	71.	GGGATTCC T AGGGGCACAGA	
		72.	GGGATTCC T CCGGGGCACAGA	
	Promoter; -1306	73.	ACAGCGCCATGTTGCTTCTGG	
		74.	ACAGCGCCAT ATT GCTTCTGG	
20	5' UTR - A; 36	75.	CAGCCACTGC GGG CGAGGAGG	
		76.	CAGCCACTGC AGG CGAGGAGG	
	5' UTR - A; 51	77.	AGGAGGGCAC AGG CCAGGTT	
		78.	AGGAGGGCAC GAG GCCAGGTT	
	5' UTR - B; 183	79.	GGGAAACATCGGGGAGAGGAG	
		80.	GGGAAACATCAGGGAGAGGAG	
25	Exon 5*; 447	81.	CCAAAGTGTACCC T CACCT	INS STOP
		82.	CCAAAGTGT A CCCTCACCT	INS STOP
	Exon 5*; 136 (nt 307)	83.	AAGGATGTGGCGTTCC T AC	ALA-VAL
		84.	AAGGATGTGGTGGTTTC T AC	ALA-VAL
30	Exon 7*; 235 (nt 705)	85.	ATGAAGAAGAAC C CTATGACC	ASN-THR
		86.	ATGAAGAAGAC CC CTATGACC	ASN-THR
	Exon 8*; 282 (nt 845)	87.	GGACTATGCGGAGAAGATGGC	GLU-LYS
35		88.	GGACTATGC G AGAAGATGGC	GLU-LYS
	Exon 2; 19 (nt 56)	89.	AAGGGGGTCCCG C TATCAAG	PRO-LEU
		90.	AAGGGGGT C CTGCTCATCAAG	PRO-LEU
	Intron 1A; 88	91.	CTCTG C TATCTCTGCC C TC	
40		92.	CTCTG C TATCC C CTGCC C TC	
	Intron 2; 34	93.	CTCT CC CAGGTGGCAGT CC CC	
		94.	CTCT CC CAGGCGGCAGT CC CC	
	Intron 4; -71	95.	CCTTG CC AA CC AAGAGATG	DEL A
		96.	CCTTG CC AC CC AAGAGATG	DEL A
45	Intron 5; -19	97.	GTGTCGGCACTCC C TGCC C G	

	98.	GTGTCGGCACCCCTGCCGC
	99.	CCTCCCTGGGCGGCCCCCTCCA
	100.	CCTCCCTGGGTGGCCCCCTCCA
	101.	TTGTTCTATGGATCCATGCTC
5	102.	TTGTTCTATGCATCCATGCTC
	103.	CATGGGCTGCTGGAGGCCTGT
	104.	CATGGGCTGCCGGAGGCCTGT
	105.	ACTGGGCCAGGACCCCTGGCA
	106.	ACTGGGCCAGAACCCCTGGCA
10	107.	CCTGCCTATCCCAGCTTCTC
	108.	CCTGCCTATCTCAGCTTCTC
	109.	GCCTATCCCATTTCTCCTC
	110.	GCCTATCCCAGCTTCTCCTC

15 Genotyping of 95 individuals from each of 3 broadly defined racial groups (African Americans, Hispanic Americans, and Caucasian Americans) for three polymorphisms produced the allele and genotype frequencies shown in Table 5.

20

25

30

35

Table 5. Allele and Genotyp Population frequencies. Polymorphism name includes gene abbreviation followed by nucleotide designation for allele 1, then nucleotide number, then nucleotide designation for allele 2. Population abbreviations are as follows: AfAm=African Americans; BAH=Caucasians from the United States; Cauc=Caucasians from California; Hisp=Hispanic from California. Allele name are designated by nucleotide and position. Genotypes are designated by position and 2 nucleotides representing each of the chromosomes in a given individual.

Locus	Calc Name	Start Position	Population	Allele Freq	Genotype Freq
STP2	STP2_Al136Val_AfAm	0	African Americans, California	STP2_Al136=0.997 STP2_Val136=0.003	STP2_136Ala/Ala=0.995 STP2_136Ala/Val=0.005 STP2_136Val/Val=0
STP2	STP2_Al136Val_Cauc2	0	Caucasian, USA	STP2_Al136=1 STP2_Val136=0	STP2_136Ala/Ala=1 STP2_136Ala/Val=0 STP2_136Val/Val=0
STP2	STP2_Al136Val_Cauc	0	Caucasians, California	STP2_Al136=1 STP2_Val136=0	STP2_136Ala/Ala=1 STP2_136Ala/Val=0 STP2_136Val/Val=0
STP2	STP2_Al136Val_Chin	0	Chinese, California	STP2_Al136=1 STP2_Val136=0	STP2_136Ala/Ala=1 STP2_136Ala/Val=0 STP2_136Val/Val=0
STP2	STP2_Al136Val_Hisp	0	Hispanics, California	STP2_Al136=1 STP2_Val136=0	STP2_136Ala/Ala=1 STP2_136Ala/Val=0 STP2_136Val/Val=0
STP2	STP2_Al136Val_Japn	0	Japanese, California	STP2_Al136=1 STP2_Val136=0	STP2_136Ala/Ala=1 STP2_136Ala/Val=0 STP2_136Val/Val=0

STP2	STP2_insA447_AfAm	0	African Americans, California	STP2_delA447=0.981 STP2_insA447=0.019	STP2_447delA/delA=0.963 STP2_447delA/insA=0.037 STP2_447insA/insA=0
STP2	STP2_insA447_Cauc2	0	Caucasian, USA	STP2_delA447=1 STP2_insA447=0	STP2_447delA/delA=1 STP2_447delA/insA=0 STP2_447insA/insA=0
STP2	STP2_insA447_Cauc	0	Caucasians, California	STP2_delA447=1 STP2_insA447=0	STP2_447delA/delA=1 STP2_447delA/insA=0 STP2_447insA/insA=0
STP2	STP2_insA447_Chin	0	Chinese, California	STP2_delA447=1 STP2_insA447=0	STP2_447delA/delA=1 STP2_447delA/insA=0 STP2_447insA/insA=0
STP2	STP2_insA447_Hisp	0	Hispanics, California	STP2_delA447=0.995 STP2_insA447=0.005	STP2_447delA/delA=0.989 STP2_447delA/insA=0.011 STP2_447insA/insA=0
STP2	STP2_insA447_Japn	0	Japanese, California	STP2_delA447=1 STP2_insA447=0	STP2_447delA/delA=1 STP2_447delA/insA=0 STP2_447insA/insA=0
STP2	STP2_Asn235Thr_AfAm	0	African Americans, California	STP2_Asn235=0.749 STP2_Thr235=0.251	STP2_235Asn/Asn=0.562 STP2_235Asn/Thr=0.373 STP2_235Thr/Thr=0.065
STP2	STP2_Asn235Thr_Cauc2	0	Caucasian, USA	STP2_Asn235=0.753 STP2_Thr235=0.247	STP2_235Asn/Asn=0.565 STP2_235Asn/Thr=0.376 STP2_235Thr/Thr=0.059

STP2	STP2_Asn235Thr_Cauc	0	Caucasians, California	STP2_Asn235=0.649 STP2_Thr235=0.351	STP2_235Asn/Asn=0.42 STP2_235Asn/Thr=0.458 STP2_235Thr/Thr=0.122
STP2	STP2_Asn235Thr_Chin	0	Chinese, California	STP2_Asn235=0.934 STP2_Thr235=0.066	STP2_235Asn/Asn=0.882 STP2_235Asn/Thr=0.105 STP2_235Thr/Thr=0.013
STP2	STP2_Asn235Thr_Hisp	0	Hispanics, California	STP2_Asn235=0.623 STP2_Thr235=0.377	STP2_235Asn/Asn=0.385 STP2_235Asn/Thr=0.476 STP2_235Thr/Thr=0.139
STP2	STP2_Asn235Thr_Japan	0	Japanese, California	STP2_Asn235=0.855 STP2_Thr235=0.145	STP2_235Asn/Asn=0.75 STP2_235Asn/Thr=0.211 STP2_235Thr/Thr=0.039
STP2	STP2_Glu282Lys_AfAm	0	African Americans, California	STP2_Glu282=0.949 STP2_Lys282=0.051	STP2_282Glu/Glu=0.899 STP2_282Glu/Lys=0.101 STP2_282Lys/Lys=0
STP2	STP2_Glu282Lys_Cauc2	0	Caucasian, USA	STP2_Glu282=1 STP2_Lys282=0	STP2_282Glu/Glu=1 STP2_282Glu/Lys=0 STP2_282Lys/Lys=0
STP2	STP2_Glu282Lys_Cauc	0	Caucasians, California	STP2_Glu282=0.997 STP2_Lys282=0.003	STP2_282Glu/Glu=0.995 STP2_282Glu/Lys=0.005 STP2_282Lys/Lys=0
STP2	STP2_Glu282Lys_Chin	0	Chinese, California	STP2_Glu282=1 STP2_Lys282=0	STP2_282Glu/Glu=1 STP2_282Glu/Lys=0 STP2_282Lys/Lys=0

STP2	STP2_Glu282Lys_Hisp	0	Hispanics, California	STP2_282Glu/Glu=0.984 STP2_282Glu/Lys=0.016 STP2_282Lys/Lys=0
STP2	STP2_Glu282Lys_Japn	0	Japanese, California	STP2_282Glu/Glu=1 STP2_282Glu/Lys=0 STP2_282Lys/Lys=0

Each of the polymorphisms identified in this study are unique and newly described. Several of the nucleotide base changes result in amino acid changes that may alter enzyme activity by any of a number of possible mechanisms. The changes in the 5' and 3' UTRs may alter regulation of transcription or transcript stability. Promoter region alterations may result 5 in altered regulation or efficiency of transcription.

All of these polymorphisms have utility. As the human genome project progresses, polymorphisms within every human gene must be identified in order to perform whole genome association studies that will be necessary for identifying genetic etiologies of complex diseases. These polymorphisms are useful for association studies.

10

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present 15 invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the 5 appended claims.

What is Claimed is:

1. An isolated nucleic acid molecule comprising a STP2 sequence polymorphism, as part of other than a naturally occurring chromosome.
- 5 2. A nucleic acid probe for detection of STP2 locus polymorphisms, comprising a polymorphic sequence listed in Table 4.
3. A nucleic acid probe according to Claim 2, wherein said probe is conjugated to a detectable marker.
- 10 4. An array of oligonucleotides comprising:
two or more probes for detection of STP2 locus polymorphisms, said probes comprising at least one form of a polymorphic sequence listed in Table 4.
- 15 5. A method for detecting in an individual a polymorphism in STP2 metabolism of a substrate, the method comprising:
analyzing the genome of said individual for the presence of at least one STP2 polymorphism listed in Table 4; wherein the presence of said predisposing polymorphism is indicative of an alteration in STP2 expression or activity.
- 20 6. A method according to Claim 5, wherein said analyzing step comprises detection of specific binding between the genomic DNA of said individual with an array of oligonucleotides comprising:
two or more probes for detection of STP2 locus polymorphisms, said probes comprising at least one form of a polymorphic sequence listed in Table 4.
- 25 7. A method according to Claim 5, wherein said alteration in STP2 expression is tissue specific.
- 30 8. A method according to Claim 5, wherein said alteration in STP2 expression is in response to a STP2 modifier.
9. A method according to Claim 8, wherein said modifier induces STP2 expression.
- 35 10. A method according to Claim 8, wherein said modifier inhibits STP2 expression.

SEQUENCE LISTING

<110> Guida, Marco
Kurth, Janice
<120> Genotyping Human Phenol Sulfotransferase
(STP2)

<130> SEQ-16P

<150> 60/088,710
<151> 1998-06-10

<160> 110

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 8396
<212> DNA
<213> H. sapiens

<400> 1

ctctccctcc	ttgtctctta	cctgcctgct	gcctggaca	ggatgaagcg	ggggcccttgc	60
gttggcccaa	ccctggctgt	tggctaagag	cccacgtat	ctgcctgtga	gaggagttcc	120
ttccggaga	accagggcag	cttctgccc	tagagggca	atgcctcagc	tgagtgcagt	180
ccccggccc	cagcctggc	cagcttggg	aagagggtgc	ccagttgtgc	aatccaggcc	240
ggggcagccg	tgtcctgatc	ttggtattca	gggctgagcc	tggagggggc	ttgtgatgcc	300
tgactctgtc	tctctctctg	gccccatgcc	ttggtagctg	tgaggcgtca	ctgctttggg	360
tgacctgatc	tggctgtat	ggatgagcac	gggggaaata	gtggaaagact	cgaattaga	420
agacgtgagt	gggctttggc	cccagcctcc	ctaccccact	ccctgtcctg	ggctgcctgt	480
gaccaaccc	gtttctgcag	gcacactgga	tagccctgct	ggagctcagt	gtccctaatac	540
ccctccagat	actggtgccc	tagggaggt	catcaaagac	cagtggaca	tcgacactag	600
cctgtttcca	cgtttcttgt	tgttttttt	tttttgtgga	gacagagttt	cactttgtt	660
gcccaggctg	gagtgcattg	gcgtgatctt	ggctcaccgc	aacctctgcc	tcccggttgc	720
aagcgtattc	cctgcctcag	cctcccaagt	agctggattt	acaggcgtgt	gccaccaggc	780
ttgactaatt	ttctatttt	agtagagaca	aggtttctcc	atgttggta	ggctggtctc	840
aaactcccg	tttcaggta	tctgcctgcc	tcggcctccc	aaagtgcgtt	gattacagg	900
gtgagccacc	gtgccaggcc	ttctccaggc	tcttggcacc	ttagccagaa	acaatttaag	960
gacaagtgc	aaagtcatga	acgtaggcag	atttcctgca	gagtaaagg	actcactgaa	1020
gaagaggaac	gtgggggtcc	tcaagagagt	gtctcatgcc	ctacaagg	tggggctgac	1080
ctttatgggc	ttcttcaact	aaagaggggt	atattcatga	agagtccagg	aaaaggtaaa	1140
gatttctcaa	gaccgtggtg	ccacaattta	cacccaaata	caggtgttcc	tggagccgtc	1200
ttggcactgg	tgggtgtacg	gtttcatatg	ttactgattt	tacagtgaga	tccttaggtga	1260
aacctacatc	aaatacagcg	ccatgttgc	tctgggtgg	cgcagccagc	ttggcctca	1320
tcctatttt	cagggactta	ttggcccttg	gcacatgcag	ctatttcaag	tttccttct	1380
ctggcactgt	gaaactgctg	cctgggattt	tctgttgc	tgctagcact	ctattaatct	1440
cacattctcg	cctctttct	gtgccacccc	ctgctggtcc	ggctgtttt	cactagagt	1500
caatacaaag	tctcagtcaa	gaggcctcc	tgaagggtgc	tgagggcagg	ggtgagacta	1560
gtagccggag	gacctgccag	tcatgggat	tcctcagggg	cacagaggag	ggaggagggg	1620
cctgtggccc	tagcagggga	gcagcctc	ctctgcctgg	aaatccatg	cctcagttt	1680
ccccgcttgc	ctctgagctc	acgcaaccct	ggaaaggctt	ggagactca	cctttactca	1740

gatgggttacctgtctc	gtgcccagct	tgaccctgga	ctttaaatag	tgaggacaaa	1800
gaacgaggag	ggtgggggga	tgcactcctt	ccacgggggc	ctgtggcttc	1860
cctccctctgg	tctctgtctg	tggagcctcc	tccaaaccca	tggaaagaaa	1920
aggggctgtg	gttcttctag	gatcttctat	cgatgttctg	tgaggtcccc	1980
gaaagctgggg	ctggctcca	gggcaatggg	actgcagtgt	ccttgttctt	2040
tggatccatg	ctctgctcca	cccctgcccc	tcaactctgc	ccacacgcat	2100
tggcctgtg	gtcagagct	ggagtgcatg	ggctgctgga	ggcctgtggg	2160
ccaggacccc	tggcaccttc	aagactggcc	tggagccagc	aggtaggtga	2220
gcctgcctat	cccagcttc	tcctccaatc	cctcccccct	cttgccctggg	2280
aaagcttgtc	ttttggagtt	caggggcagg	tcaggagccc	agtgcacagct	2340
accccaaaaa	aaaaaccca	ccattggcc	cttcccccctt	tcattttctt	2400
caccaaaccc	agtcgtggct	ttggagatca	ctttaagctt	gtctccagct	2460
ggagggtaat	agagaagctc	ccccaccccc	aaccctaccc	cttccttccg	2520
taagtcacg	cccggtctca	gatccctccc	acactgacact	aagaaaccc	2580
aacacccctg	cattcccccac	acaacaccca	cactcagcoa	ctgcgggcga	2640
aggccaggtt	cccaagagct	caggtgagtg	acacacccgga	atggcccagg	2700
cctgctcagc	tttgtggctcc	aacattccag	aagccgagggc	ctctgctatc	2760
ccccatggat	atcccatttc	agacaacccc	ggccggcctg	aatccccctc	2820
ttttttccg	gggaggccag	gtcttgctgt	caccgagget	ggagtgtgt	2880
ccactgcage	cttgaattcc	tgggtcaag	tgattctcct	gcctcagtag	2940
agaccctcac	cacccctgcct	ggatagttt	aaaaaatatt	tttaaaagat	3000
ggggcttccc	aatgctgccc	agatggtct	ccaaattctg	gcctcagcct	3060
tgggattaca	ggtgggagcc	accctgcccc	ggatectcct	tttgcgtgat	3120
ttgctcatc	ccacatcagg	ctctggcccc	caataccagc	tcagttgtctc	3180
ttgtcctgga	acccagatgg	actgtggccg	ggcaagtggg	tcacaggcct	3240
ggagttgcca	catgtgaggg	gccgaggggc	tcaaggaggg	gaacatcggg	3300
tactgggtgg	aggctggggg	tcccagcagg	aatggtgag	acaaagggcg	3360
ggaagacagc	acaggaaggt	cctagaggtt	cctcagtgca	gctggactct	3420
ttcacacacc	ctgacatctg	ggcccccgttc	cacgaggggtg	ctttcactgg	3480
ggcccaggcc	ctgggatttt	gaacagctcc	gcaggtgaat	gaaaggtgag	3540
ggaaccacca	cattagaacc	cgacctggtt	ttagccccca	gccccccac	3600
tgtgagtgcg	ggcaagtac	tcaacctccc	taggcctcag	tgactccct	3660
atccacttt	tttgctgttg	tgatggtggt	aagggaaacgg	gcctggctct	3720
gcaggaacat	ggagctgatc	caggacatct	ctcgccccgc	actggagtagc	3780
tcccgctcat	caagtacttt	gcagaggcac	tggggccccc	gcaagacttc	3840
ctgatgacct	gctcatcagc	acctacccca	agtccggtag	gtgaggaggg	3900
tctccctagg	ggcagtcctcc	accttggcca	gcgaggtcat	gctcacctca	3960
ctcccatctc	cctccctctc	caggaccac	ctgggtgagc	cagattctgg	4020
ccagggcggt	gacctggaaa	agtgtcaccg	agctcccatc	ttcatgcggg	4080
tgagttcaaa	gtcccctcagg	tgtgtgtgtc	ctgggtgcaa	ggggagtgga	4140
ggaagacagg	gctggggctt	cagtcacca	gaccttccct	gacccactgc	4200
agactctgaa	aaacacacca	gccccacgac	tcctgaagac	acacctgccc	4260
tcccccagac	tctgttggat	cagaaggctca	aggtgagact	gggcacagtg	4320
gcaatctcag	tactttggga	ggctgaggtg	ggaagatccc	ttgaagccag	4380
taagtctctt	ccaaaaaaaaa	aacttagctg	tgcatagtgg	tgtgtgcctg	4440
tactcaggag	gttgagggtgg	gaggatcatc	tgacgcctagg	agtttaaggt	4500
tatgatcaca	ccagtgcact	ccaggtgggg	tgacagagaa	acactgtctc	4560
gaatagaaag	agtgtcccac	cagtgcgggt	gctcacacct	gtaattccag	4620
ggctgaggca	ggtggtacac	ctgagactag	gagtttgaga	tcagcctggc	4680
aaaccccatc	tctactaaaa	atacaaaaaa	attagccggg	catggtggca	4740
atcccagcta	cttgggaggc	tgaagcagga	gaattgtctg	aagctgggag	4800
tagtcagccg	agacctcacc	attgcaccgc	agcctggaa	acaagagcaa	4860
aaaaaaaaa	agaaaaaaaaa	aaaaaagcgg	caggtggcag	ggggctggc	4920

tcacgcctgt aataccagca	cttcggagg tcgaggtggg	cagatcaccc aaggttagga	4980
gtttgagatc agtctggcca	acatggagaa accccgtctc	tactaaaaat aaaaaatta	5040
gccaggcggt ggggcaggcg	ccagtaatcc cagctactcg	ggaggctgag gaaggagaat	5100
agcttgcacc tgggaggcg	tggttgcagt gagccagat	tgtgccactg tactccagcc	5160
tgggagacac aacgagacat	tgtttcaaac aaaacaaata	aatattttaa aaggtttgcc	5220
acctgggtgg ctcaccgctg	taatgccagc atttggag	gccaagatgg gtggaccgct	5280
tgagctcagg agttccagac	cagcccagga aacatggga	gactccatct ctataaaaaga	5340
tgcaaataat cagcagggca	tggtggcata gcgcata	cccagctact caaaagtcta	5400
aggttggagg attgcttgc	cctgggaggt caacgttgca	gtgagctatt ctcaactccag	5460
tgcactccaa cctggcaac	aggaaaaaaag aaagcccaag	gtcttttcc tctttctct	5520
ttttttttag acctagagtc	cccccccca aaaaaaaaaa	aaccacaaca aaaaagaaaaa	5580
agcaaaggc caggtgtgg	gcatgtgaat ccagggagg	aggccccggc tcagcccagc	5640
tttggcctg ttcttctggg	agagtcgcct cacttcctcc	agacttgtct catttccac	5700
gggggggact gtctgcctt	tgctctgatg accaaaaaca	tgagactctt ccggtagac	5760
ctaagaaagg tagagggtag	gtcctcacag acccacaaaa	tttgggtggt gtggaaacat	5820
gcctggtga gcatgcctt	ctccagatcg gggtgtgacg	cattgatgca gattatatta	5880
ctatagaata tgatggctc	agggaccagg caggacttt	gtcttgagc agggttcaga	5940
tcctgacttgc cccctacctg	tgccgtgaga tctcaaaca	gtcagcctct aagcctcagc	6000
ttcctccctt gccaaaccaa	gagatgagct ggctggggc	aggctgtgt gtgatggtag	6060
tggggttgag tcttctgc	ctgcaggtgg tctatgttgc	ccgcaacgca aaggatgtgg	6120
cggtttccat ctaccactt	taccacatgg ccaaagtgt	ccctcacccct gggacctggg	6180
aaagcttccct ggagaagttc	atggctggag aaggtgggct	tgatgggagg aaggaaggtag	6240
tggagctaag ggggtggc	tacaacgcac agcaaccctg	tgtcggacc cccgtccgc	6300
ttctccatgt tcctatgggt	cctggatcca gcacgtgca	gagtgggtgg agctgagccg	6360
caccacccct gttctctacc	tcttctatga agacatgaag	gaggtgagac cgcctttgat	6420
gcttccctcc acgtgacacc	tggggcagg cacttcacag	ggacctgcca aggccaccca	6480
gccaccctcc ctggggggcc	cctccagcag gcccggattc	cccatctgta ctcccggcc	6540
caggccccac tgcagcccca	tgtggcagca ggctgggcac	agctctcata tcctgtgcct	6600
gagtcagctg cacgggtggc	catggatcag ctacttttt	ttttgagaca aaagtcttgc	6660
tctttgtcc aggatggcat	gcagttgtgt gatctcagct	cagtgtaaacc cccctcccca	6720
ggttcaagtg attctctgc	tcagcctcc tgagtagctg	agattacaga tgacactac	6780
catgcctggc taatttttgt	gttgcctat gttggccagg	ttggctc tctcctgagc	6840
tcaggtgatc cgcctgcctc	agcctccaa agtcttggga	attacacgca tgaaccacgg	6900
ccccctgcca cagatcagct	atctattcca attgttctc	cctgccaatg gttatggcac	6960
ccagggccac aggacacggaa	ccagtcctta cccataggag	ccaagcccaag	7020
ctcatgatgg gatcacagg	cagacagcaa ttcatttgc	cccaggact ggggtcccaag	7080
gggtcgagga gtcggctcta	tgggtttga agtggaaatg	gccagttccc ctctgaggtt	7140
agagaagtgg acccccttta	ttttctgaa tcagcaatcc	aagcctccac tgaggagccc	7200
tctgtgtcc agaacccttca	aaggagatt caaaagatcc	tggagttgtt gggcgctcc	7260
ctgcacagg agactgtgga	cctcatggtt gacacacgt	cggtcaagga gatgaagaag	7320
aaccctatga ccaactacac	caccgtccgc cgggagttca	tggaccacag catctccccc	7380
ttcatgagga aagtaggtg	ccggccagca cgggggtttt	gagcaggtgg gagcagcagc	7440
tggagcctcc ccataggcac	tcggggcctc ccctggatg	agactccagc ttgtccct	7500
gccttcctcc cccaggcatg	gctggggact ggaagaccac	cttcacccgt ggcagaatg	7560
agcgttgcga tgcggactat	gcccagaaga tggcaggctg	cagcctcagc ttccgtctg	7620
agctgtgaga ggggttctg	gagtcactgc agagggagtg	tgcaatcaa gcctgaccac	7680
gaggctccag aataaaagtat	gatttgcgtt caatgcagag	tctctattcc aagccaagag	7740
aaaccctgag ctgaaagagt	gatgcacccac tggggccaaa	tacggccacc tccccgctcc	7800
agctcctcaa ttgcctgtt	ttggagaggg gagagggtct	ggagaagtaa aaccaggag	7860
acgagtagag gggaaatgtg	tttaatccca gcacgtcctc	tgctgtctg ccctgtgtcg	7920
ttggggatg gcgagtc	caggccgcat cacttttct	tgggttcctt acaagccacc	7980
acgtatctct gagccacatt	gaggggaggg gaatagccat	ctgcata tagga ggtgtctca	8040
aacaggaccg agtagtc	ctggggctgtt gggcaggca	gacaggaggg gctgctcaga	8100

gacccccagg ccaggacagg cacccttc ccccagcta gaccacagga ggctctggc	8160
cgtggactct cagccactcc taacatcctt cactctgggg tcaagaagtc ttggcccagt	8220
ccctgctgt acagagctct ttctcagtg gctggagacc caaggcaggg aataggcagg	8280
gaggagtagg ggtgctgact cccttcctag tgggtcata gctggagggt ctgtgcctt	8340
tcaaggactc ttgttgaga ggactgaggg caacccagag gggtggcaggc agggat	8396

<210> 2
 <211> 1396
 <212> DNA
 <213> *H. sapiens*

<220>
 <221> CDS
 <222> (426) ... (1308)

<400> 2				
gcattccccca cacaacaccc acactcagcc actgcgggac aggagggcac gaggccaggt	60			
tcccaagagc tcaggttgt cctggAACCC agatggactg tggccggca agtggatcac	120			
aggcctggcc agcctaggag ttgccacatg tgaggggccc aggggctcaa ggaggggaac	180			
atcggggaga ggagcctact gggtgaggc tgggggtccc agcaggaaat ggtgagacaa	240			
agggcgctgg ctggcagggaa gacagcacag gaaggtccta gagtttctc agtgcagctg	300			
gactcttcctg gagaccttca cacaccctga catctggcc ccttccacg agggtgcctt	360			
cactggtctg caccatggcc caggccctgg gatTTGAAC agtccgcag gtgaatgaaa	420			
ggaac atg gag ctg atc cag gac atc tct cgc ccg cca ctg gag tac gtg	470			
Met Glu Leu Ile Gln Asp Ile Ser Arg Pro Pro Leu Glu Tyr Val				
1	5	10	15	
aag ggg gtc ccg ctc atc aag tac ttt gca gag gca ctg ggg ccc ctg				518
Lys Gly Val Pro Leu Ile Lys Tyr Phe Ala Glu Ala Leu Gly Pro Leu				
20	25	30		
cag agc ttc cag gcc cgg cct gat gac ctg ctc atc agc acc tac ccc				566
Gln Ser Phe Gln Ala Arg Pro Asp Asp Leu Leu Ile Ser Thr Tyr Pro				
35	40	45		
aag tcc ggc acc acc tgg gtg agc cag att ctg gac atg atc tac cag				614
Lys Ser Gly Thr Thr Trp Val Ser Gln Ile Leu Asp Met Ile Tyr Gln				
50	55	60		
ggc ggt gac ctg gaa aag tgg cac cga gct ccc atc ttc atg cgg gtg				662
Gly Gly Asp Leu Glu Lys Cys His Arg Ala Pro Ile Phe Met Arg Val				
65	70	75		
ccc ttc ctt gag ttc aaa gtc cca ggg att ccc tca ggg atg gag act				710
Pro Phe Leu Glu Phe Lys Val Pro Gly Ile Pro Ser Gly Met Glu Thr				
80	85	90	95	
ctg aaa aac aca cca gcc cca cga ctc ctg aag aca cac ctg ccc ctg				758
Leu Lys Asn Thr Pro Ala Pro Arg Leu Leu Lys Thr His Leu Pro Leu				
100	105	110		

gct ctg ctc ccc cag act ctg ttg gat cag aag gtc aag gtg gtc tat Ala Leu Leu Pro Gln Thr Leu Leu Asp Gln Lys Val Lys Val Val Tyr 115 120 125	806
gtt gcc cgc aac gca aag gat gtg gcg gtt tcc tac tac cac ttc tac Val Ala Arg Asn Ala Lys Asp Val Ala Val Ser Tyr Tyr His Phe Tyr 130 135 140	854
cac atg gcc aaa gtg tac cct cac cct ggg acc tgg gaa agc ttc ctg His Met Ala Lys Val Tyr Pro His Pro Gly Thr Trp Glu Ser Phe Leu 145 150 155	902
gag aag ttc atg gct gga gaa gtg tcc tat ggg tcc tgg tac cag cac Glu Lys Phe Met Ala Gly Glu Val Ser Tyr Gly Ser Trp Tyr Gln His 160 165 170 175	950
gtg caa gag tgg tgg gag ctg agc cgc acc cac cct gtt ctc tac ctc Val Gln Glu Trp Trp Glu Leu Ser Arg Thr His Pro Val Leu Tyr Leu 180 185 190	998
ttc tat gaa gac atg aag gag aac ccc aaa agg gag att caa aag atc Phe Tyr Glu Asp Met Lys Glu Asn Pro Lys Arg Glu Ile Gln Lys Ile 195 200 205	1046
ctg gag ttt gtg ggg cgc tcc ctg cca gag gag act gtg gac ctc atg Leu Glu Phe Val Gly Arg Ser Leu Pro Glu Glu Thr Val Asp Leu Met 210 215 220	1094
gtt gag cac acg tcg ttc aag gag atg aag aac cct atg acc aac Val Glu His Thr Ser Phe Lys Glu Met Lys Lys Asn Pro Met Thr Asn 225 230 235	1142
tac acc acc gtc cgc cgg gag ttc atg gac cac agc atc tcc ccc ttc Tyr Thr Val Arg Arg Glu Phe Met Asp His Ser Ile Ser Pro Phe 240 245 250 255	1190
atg agg aaa ggc atg gct ggg gac tgg aag acc acc ttc acc gtg gcg Met Arg Lys Gly Met Ala Gly Asp Trp Lys Thr Thr Phe Thr Val Ala 260 265 270	1238
cag aat gag cgc ttc gat gcg gac tat gcg gag aag atg gca ggc tgc Gln Asn Glu Arg Phe Asp Ala Asp Tyr Ala Glu Lys Met Ala Gly Cys 275 280 285	1286
agc ctc agc ttc cgc tct gag c tgtgagaggg gttccctggag tcactgcaga Ser Leu Ser Phe Arg Ser Glu 290	1338
gggagtgtgc gaatcaagcc tgaccaagag gctccagaat aaagtatgtat ttgtgttc	1396
<210> 3	
<211> 295	
<212> PRT	

<213> H. sapiens

<400> 3

Met	Glu	Leu	Ile	Gln	Asp	Ile	Ser	Arg	Pro	Pro	Leu	Glu	Tyr	Val	Lys
1						5			10						15
Gly	Val	Pro	Leu	Ile	Lys	Tyr	Phe	Ala	Glu	Ala	Leu	Gly	Pro	Leu	Gln
						20			25					30	
Ser	Phe	Gln	Ala	Arg	Pro	Asp	Asp	Leu	Leu	Ile	Ser	Thr	Tyr	Pro	Lys
						35			40				45		
Ser	Gly	Thr	Thr	Trp	Val	Ser	Gln	Ile	Leu	Asp	Met	Ile	Tyr	Gln	Gly
						50			55			60			
Gly	Asp	Leu	Glu	Lys	Cys	His	Arg	Ala	Pro	Ile	Phe	Met	Arg	Val	Pro
						65			70			75			80
Phe	Leu	Glu	Phe	Lys	Val	Pro	Gly	Ile	Pro	Ser	Gly	Met	Glu	Thr	Leu
						85			90			95			
Lys	Asn	Thr	Pro	Ala	Pro	Arg	Leu	Leu	Lys	Thr	His	Leu	Pro	Leu	Ala
						100			105			110			
Leu	Leu	Pro	Gln	Thr	Leu	Leu	Asp	Gln	Lys	Val	Lys	Val	Val	Tyr	Val
						115			120			125			
Ala	Arg	Asn	Ala	Lys	Asp	Val	Ala	Val	Ser	Tyr	Tyr	His	Phe	Tyr	His
						130			135			140			
Met	Ala	Lys	Val	Tyr	Pro	His	Pro	Gly	Thr	Trp	Glu	Ser	Phe	Leu	Glu
						145			150			155			160
Lys	Phe	Met	Ala	Gly	Glu	Val	Ser	Tyr	Gly	Ser	Trp	Tyr	Gln	His	Val
						165			170			175			
Gln	Glu	Trp	Trp	Glu	Leu	Ser	Arg	Thr	His	Pro	Val	Leu	Tyr	Leu	Phe
						180			185			190			
Tyr	Glu	Asp	Met	Lys	Glu	Asn	Pro	Lys	Arg	Glu	Ile	Gln	Lys	Ile	Leu
						195			200			205			
Glu	Phe	Val	Gly	Arg	Ser	Leu	Pro	Glu	Glu	Thr	Val	Asp	Leu	Met	Val
						210			215			220			
Glu	His	Thr	Ser	Phe	Lys	Glu	Met	Lys	Lys	Asn	Pro	Met	Thr	Asn	Tyr
						225			230			235			240
Thr	Thr	Val	Arg	Arg	Glu	Phe	Met	Asp	His	Ser	Ile	Ser	Pro	Phe	Met
						245			250			255			
Arg	Lys	Gly	Met	Ala	Gly	Asp	Trp	Lys	Thr	Thr	Phe	Thr	Val	Ala	Gln
						260			265			270			
Asn	Glu	Arg	Phe	Asp	Ala	Asp	Tyr	Ala	Glu	Lys	Met	Ala	Gly	Cys	Ser
						275			280			285			
Leu	Ser	Phe	Arg	Ser	Glu	Leu									
						290			295						

<210> 4

<211> 18

<212> DNA

<213> H. sapiens

<400> 4

cccaaataca ggtgttcc

18

<210> 5

<211> 17

<212> DNA

<213> H. sapiens
<400> 5
ggagcagagc aaggatc 17
<210> 6
<211> 21
<212> DNA
<213> H. sapiens
<400> 6
ttcttctagg atcttctatc g 21
<210> 7
<211> 18
<212> DNA
<213> H. sapiens
<400> 7
actcagcaaa aggaggat 18
<210> 8
<211> 18
<212> DNA
<213> H. sapiens
<400> 8
tttagagatgg ggtcttcc 18
<210> 9
<211> 15
<212> DNA
<213> H. sapiens
<400> 9
gggcgagaga tgtcc 15
<210> 10
<211> 17
<212> DNA
<213> H. sapiens
<400> 10
ggagaggagc ctactgg 17
<210> 11
<211> 16
<212> DNA
<213> H. sapiens
<400> 11
agtctgaggt gagcat 16

<210> 12
<211> 17
<212> DNA
<213> H. sapiens

<400> 12
gcctcagtga cttccct 17

<210> 13
<211> 20
<212> DNA
<213> H. sapiens

<400> 13
tttggaaagag acttatctgg 20

<210> 14
<211> 16
<212> DNA
<213> H. sapiens

<400> 14
gcaggacttt ggcttt 16

<210> 15
<211> 16
<212> DNA
<213> H. sapiens

<400> 15
gactcaggca caggag 16

<210> 16
<211> 16
<212> DNA
<213> H. sapiens

<400> 16
gaccatccca gtcctt 16

<210> 17
<211> 15
<212> DNA
<213> H. sapiens

<400> 17
ccccaacgac acagg 15

<210> 18
<211> 15
<212> DNA
<213> H. sapiens

<400> 18
tggagcccg tttgg 15

<210> 19
<211> 18
<212> DNA
<213> H. sapiens

<400> 19
cagcagtttc acttgacc 18

<210> 20
<211> 14
<212> DNA
<213> H. sapiens

<400> 20
tgccacccccc tgct 14

<210> 21
<211> 14
<212> DNA
<213> H. sapiens

<400> 21
aggctgctcc cctg 14

<210> 22
<211> 14
<212> DNA
<213> H. sapiens

<400> 22
gggctcacgc aacc 14

<210> 23
<211> 19
<212> DNA
<213> H. sapiens

<400> 23
gcaggtactt ttctttcca 19

<210> 24
<211> 21
<212> DNA
<213> H. sapiens

<400> 24
ttcttctagg atcttctatc g 21

<210> 25
<211> 18

<212> DNA
<213> H. sapiens

<400> 25
ttttttaggt gtcactgg 18

<210> 26
<211> 16
<212> DNA
<213> H. sapiens

<400> 26
cccacacacaac acccac 16

<210> 27
<211> 16
<212> DNA
<213> H. sapiens

<400> 27
gcttctggaa tgttgg 16

<210> 28
<211> 19
<212> DNA
<213> H. sapiens

<400> 28
cgaaaaaaa aaaaggaag 19

<210> 29
<211> 14
<212> DNA
<213> H. sapiens

<400> 29
caatgctgcc caga 14

<210> 30
<211> 17
<212> DNA
<213> H. sapiens

<400> 30
gctccactga ggaacct 17

<210> 31
<211> 17
<212> DNA
<213> H. sapiens

<400> 31
ggagagggagc ctactgg 17

<210> 32
<211> 18
<212> DNA
<213> H. sapiens

<400> 32
taccaccatc acaacagc 18

<210> 33
<211> 19
<212> DNA
<213> H. sapiens

<400> 33
ctgaaagcaa gaaatccac 19

<210> 34
<211> 16
<212> DNA
<213> H. sapiens

<400> 34
aggctgaggt gagcat 16

<210> 35
<211> 14
<212> DNA
<213> H. sapiens

<400> 35
gcggtgacct ggaa 14

<210> 36
<211> 20
<212> DNA
<213> H. sapiens

<400> 36
tttggaaagag acttatctgg 20

<210> 37
<211> 17
<212> DNA
<213> H. sapiens

<400> 37
ctgacttgcc cctacct 17

<210> 38
<211> 16
<212> DNA
<213> H. sapiens

<400> 38	
tagccaccac ccctta	16
<210> 39	
<211> 18	
<212> DNA	
<213> H. sapiens	
<400> 39	
ccaaagtgtta ccctcacc	18
<210> 40	
<211> 14	
<212> DNA	
<213> H. sapiens	
<400> 40	
agcctgctgc caca	14
<210> 41	
<211> 16	
<212> DNA	
<213> H. sapiens	
<400> 41	
gaccatccca gtcctt	16
<210> 42	
<211> 14	
<212> DNA	
<213> H. sapiens	
<400> 42	
caaaccggcg tgct	14
<210> 43	
<211> 18	
<212> DNA	
<213> H. sapiens	
<400> 43	
ctgtggacct cttggttt	18
<210> 44	
<211> 22	
<212> DNA	
<213> H. sapiens	
<400> 44	
cacaaatcat actttattct gg	22
<210> 45	

<211> 15
<212> DNA
<213> H. sapiens

<400> 45
cgatgcggac tatgc 15

<210> 46
<211> 15
<212> DNA
<213> H. sapiens

<400> 46
ccccaacgac acagg 15

<210> 47
<211> 22
<212> DNA
<213> H. sapiens

<400> 47
ggtgctgggg ttgagtcttc tg 22

<210> 48
<211> 27
<212> DNA
<213> H. sapiens

<400> 48
caaaggatgt ggccggttcc tactacc 27

<210> 49
<211> 23
<212> DNA
<213> H. sapiens

<400> 49
acacaccttccatc aag 23

<210> 50
<211> 28
<212> DNA
<213> H. sapiens

<400> 50
cgcaaaggat gtgggtggttt cctactac 28

<210> 51
<211> 24
<212> DNA
<213> H. sapiens

<400> 51

ggagactgtg gacccatgg ttga	24
<210> 52	
<211> 32	
<212> DNA	
<213> H. sapiens	
<400> 52	
tagttggtca tagggttctt cttcatctcc tt	32
<210> 53	
<211> 21	
<212> DNA	
<213> H. sapiens	
<400> 53	
ccggcaccta cctttcctca t	21
<210> 54	
<211> 30	
<212> DNA	
<213> H. sapiens	
<400> 54	
tagttggtca taggggtctt cttcatctcc	30
<210> 55	
<211> 21	
<212> DNA	
<213> H. sapiens	
<400> 55	
agctttgctc cctgccttcc t	21
<210> 56	
<211> 24	
<212> DNA	
<213> H. sapiens	
<400> 56	
ctgccatctt ctccgcatacg tccg	24
<210> 57	
<211> 22	
<212> DNA	
<213> H. sapiens	
<400> 57	
ggaacccctc tcacagctca ga	22
<210> 58	
<211> 24	
<212> DNA	

<213> H. sapiens

<400> 58
tgccatcttc ttccatgtt ccgc 24

<210> 59
<211> 22
<212> DNA
<213> H. sapiens

<400> 59
ggtgctgggg ttgagtcttc tg 22

<210> 60
<211> 25
<212> DNA
<213> H. sapiens

<400> 60
atggccaaag tgtaccctca ccctg 25

<210> 61
<211> 23
<212> DNA
<213> H. sapiens

<400> 61
acaccccttcc tcctccatc aag 23

<210> 62
<211> 25
<212> DNA
<213> H. sapiens

<400> 62
catggccaaa gtgttaaccct caccc 25

<210> 63
<211> 21
<212> DNA
<213> H. sapiens

<400> 63
ccagctccctc aacttgccct g 21

<210> 64
<211> 21
<212> DNA
<213> H. sapiens

<400> 64
ccagctccctc tacttgccct g 21

<210> 65
<211> 21
<212> DNA
<213> H. sapiens

<400> 65
gtgagagggg ttcctggagt c 21

<210> 66
<211> 21
<212> DNA
<213> H. sapiens

<400> 66
gtgagagggg ctcctggagt c 21

<210> 67
<211> 21
<212> DNA
<213> H. sapiens

<400> 67
catgaagctg gggctggctc c 21

<210> 68
<211> 21
<212> DNA
<213> H. sapiens

<400> 68
catgaagctg aggctggctc c 21

<210> 69
<211> 21
<212> DNA
<213> H. sapiens

<400> 69
ctcgtgcccc a cttgaccct g 21

<210> 70
<211> 21
<212> DNA
<213> H. sapiens

<400> 70
ctcgtgcccc a cttgaccct g 21

<210> 71
<211> 21
<212> DNA
<213> H. sapiens

<400> 71
gggatttcctc agggggcacag a 21

<210> 72
<211> 21
<212> DNA
<213> H. sapiens

<400> 72
gggatttcctc cggggcacag a 21

<210> 73
<211> 21
<212> DNA
<213> H. sapiens

<400> 73
acagcgccat gttgcttctg g 21

<210> 74
<211> 21
<212> DNA
<213> H. sapiens

<400> 74
acagcgccat attgcttctg g 21

<210> 75
<211> 21
<212> DNA
<213> H. sapiens

<400> 75
cagccactgc gggcgaggag g 21

<210> 76
<211> 21
<212> DNA
<213> H. sapiens

<400> 76
cagccactgc aggcgaggag g 21

<210> 77
<211> 21
<212> DNA
<213> H. sapiens

<400> 77
aggagggcac aaggccaggt t 21

<210> 78
<211> 21

<212> DNA
<213> H. sapiens

<400> 78
aggagggcac gaggccaggt t 21

<210> 79
<211> 21
<212> DNA
<213> H. sapiens

<400> 79
ggggAACATC ggggagagga g 21

<210> 80
<211> 21
<212> DNA
<213> H. sapiens

<400> 80
ggggAACATC agggagagga g 21

<210> 81
<211> 20
<212> DNA
<213> H. sapiens

<400> 81
ccaaAGTgta ccctCACCT 20

<210> 82
<211> 21
<212> DNA
<213> H. sapiens

<400> 82
ccaaAGTgta accCTCACCC t 21

<210> 83
<211> 20
<212> DNA
<213> H. sapiens

<400> 83
aaggATgtgg ggTTTcCTAC 20

<210> 84
<211> 20
<212> DNA
<213> H. sapiens

<400> 84
aaggATgtgg ggTTTcCTAC 20

<210> 85
<211> 20
<212> DNA
<213> H. sapiens

<400> 85
atgaagaaga ccctatgacc 20

<210> 86
<211> 20
<212> DNA
<213> H. sapiens

<400> 86
atgaagaaga ccctatgacc 20

<210> 87
<211> 20
<212> DNA
<213> H. sapiens

<400> 87
ggactatgcg agaagatggc 20

<210> 88
<211> 20
<212> DNA
<213> H. sapiens

<400> 88
ggactatgcg agaagatggc 20

<210> 89
<211> 21
<212> DNA
<213> H. sapiens

<400> 89
aagggggtcc cgctcatcaa g 21

<210> 90
<211> 21
<212> DNA
<213> H. sapiens

<400> 90
aagggggtcc tgctcatcaa g 21

<210> 91
<211> 21
<212> DNA
<213> H. sapiens

<400> 91	
ctctgctatc tctgccctct c	21
<210> 92	
<211> 21	
<212> DNA	
<213> H. sapiens	
<400> 92	
ctctgctatc cctgccctct c	21
<210> 93	
<211> 21	
<212> DNA	
<213> H. sapiens	
<400> 93	
ctctcccagg tggcagtccc c	21
<210> 94	
<211> 21	
<212> DNA	
<213> H. sapiens	
<400> 94	
ctctcccagg cggcagtccc c	21
<210> 95	
<211> 20	
<212> DNA	
<213> H. sapiens	
<400> 95	
cctttgccaa ccaagagatg	20
<210> 96	
<211> 19	
<212> DNA	
<213> H. sapiens	
<400> 96	
cctttgccac caagagatg	19
<210> 97	
<211> 21	
<212> DNA	
<213> H. sapiens	
<400> 97	
gtgtcgac tccctgcccc c	21
<210> 98	

<211> 21
<212> DNA
<213> H. sapiens

<400> 98
gtgtcgac cccctgcccc c 21

<210> 99
<211> 21
<212> DNA
<213> H. sapiens

<400> 99
cctccctggg cggccccctcc a 21

<210> 100
<211> 21
<212> DNA
<213> H. sapiens

<400> 100
cctccctggg tggccccctcc a 21

<210> 101
<211> 21
<212> DNA
<213> H. sapiens

<400> 101
ttgttctatg gatccatgtc c 21

<210> 102
<211> 21
<212> DNA
<213> H. sapiens

<400> 102
ttgttctatg catccatgtc c 21

<210> 103
<211> 21
<212> DNA
<213> H. sapiens

<400> 103
catgggctgc tggaggcctg t 21

<210> 104
<211> 21
<212> DNA
<213> H. sapiens

<400> 104

catgggctgc cgaggaggctg t 21
<210> 105
<211> 21
<212> DNA
<213> H. sapiens

<400> 105
actgggccag gaccctggc a 21

<210> 106
<211> 21
<212> DNA
<213> H. sapiens

<400> 106
actgggccag aaccctggc a 21

<210> 107
<211> 21
<212> DNA
<213> H. sapiens

<400> 107
cctgcctatac ccagctttct c 21

<210> 108
<211> 21
<212> DNA
<213> H. sapiens

<400> 108
cctgcctatac tcagctttct c 21

<210> 109
<211> 21
<212> DNA
<213> H. sapiens

<400> 109
gcctatccca tctttctcct c 21

<210> 110
<211> 21
<212> DNA
<213> H. sapiens

<400> 110
gcctatccca gctttctcct c 21

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/13094

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HER C ET AL: "Human phenol sulfotransferase STP2 gene: molecular cloning, structural characterization, and chromosomal localization." GENOMICS, (1996 MAY 1) 33 (3) 409-20. , XP002117271 cited in the application the whole document ---	1-4
X	GAEDIGK: "Cloning structural organization, and chromosomal mapping of human phenol sulfotransferase STP2 gene" GENOMICS, vol. 40, 1997, pages 242-246, XP002117277 the whole document --- -/-	1



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 October 1999	15/10/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016	Authorized officer Reuter, U

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/13094

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DOOLEY T P ET AL: "Genomic organization and DNA sequences of two human phenol sulfotransferase genes (STP1 and STP2) on the short arm of chromosome 16." BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, (1996 NOV 1) 228 (1) 134-40. , XP002117273 the whole document ----	1
A	WEINSHILBOUM R M ET AL: "Sulfation and sulfotransferases 1: Sulfotransferase molecular biology: cDNAs and genes." FASEB JOURNAL, (1997 JAN) 11 (1) 3-14. REF: 99 , XP002117274 the whole document -----	1-10